

Câu 1: (1 điểm) Cho X, Y là các biến ngẫu nhiên rời rạc có phân phối xác suất đồng thời như sau:

$X \backslash Y$	-3	2	4
1	0,1	0,2	0,2
3	0,3	0,1	0,1

a. Hai biến ngẫu nhiên X và Y có độc lập không? Vì sao?

b. Tính $P(X > 2 | Y > 1)$.

Câu 2: (1,5 điểm) Cho hai biến ngẫu nhiên X, Y có hàm mật độ xác suất đồng thời

$$f(x, y) = \begin{cases} \frac{3}{16}xy^2, & 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm hàm mật độ xác suất thành phần của X .

b. Tính $P(Y \leq 1 | X = 1)$.

c. Tính $P(X \leq Y)$.

Câu 3: (1 điểm) Để tìm hiểu thời gian xe chạy (tính bằng phút) của các xe (xe máy hoặc ô tô) chạy từ cầu Hóa An (Biên Hòa) đến cầu Nguyễn Văn Cừ (Quận 5, TPHCM), người ta thu được dữ liệu dưới đây

Thời gian	(40,45]	(45,50]	(50,55]	(55,60]	(60,65]	(65,70]	(70,75]
Số xe	4	14	23	36	17	10	6

Tính trung bình và độ lệch chuẩn của thời gian xe chạy trong mẫu trên.

Câu 4: (3 điểm) Khảo sát 750 sinh viên ngành Khoa học máy tính, người ta thấy rằng có 300 sinh viên thường xuyên sử dụng trí tuệ nhân tạo (AI) để hỗ trợ làm bài tập.

a. Hãy ước lượng tỉ lệ sinh viên ngành Khoa học máy tính sử dụng AI để làm bài tập với độ tin cậy 90%.

b. Nếu muốn sai số ước lượng không quá 2% thì cần ít nhất bao nhiêu sinh viên trong mẫu khảo sát với độ tin cậy 95%?

c. Một giảng viên cho rằng tỉ lệ sinh viên sử dụng AI để làm bài tập nhiều hơn 38%. Với dữ kiện khảo sát bên trên và mức ý nghĩa 5%, ta có đủ bằng chứng để ủng hộ phát biểu của giảng viên đó không?

Câu 5: (2,5 điểm) Một công ty quản lý hệ thống máy chủ (server) tin rằng thời gian phản hồi trung bình của các máy chủ đã được cải thiện so với năm 2020. Vào năm 2020, thời gian phản hồi trung bình của máy chủ là 250 ms (mili giây). Để kiểm chứng điều này, nhóm kỹ thuật đã thu thập dữ liệu từ 20 máy chủ ngẫu nhiên trong hệ thống và đo thời gian phản hồi (tính bằng mili giây). Người ta thấy rằng thời gian phản hồi trung bình của mẫu này là 242,92 ms và độ lệch chuẩn của mẫu là 10,39 ms.

- Hãy ước lượng thời gian phản hồi trung bình các máy chủ với độ tin cậy 95%.
- Sử dụng mức ý nghĩa 5%, ta có đủ bằng chứng để ủng hộ cho tuyên bố thời gian phản hồi trung bình của máy chủ đã giảm so với năm 2020 không? Giả sử thời gian phản hồi tuân theo phân phối chuẩn.

Câu 6. (1 điểm) Một công ty phát triển phần mềm muốn dự đoán thời gian hoàn thành dự án (tính bằng ngày) dựa trên số lượng dòng code (LOC - Lines of Code). Dữ liệu thu thập được từ 10 dự án trước đó như sau (X là số dòng code, Y là số ngày hoàn thành dự án):

X	1200	1500	1700	2100	2500	3000	3200	3600	4000	4400
Y	15	18	19	23	28	32	35	38	42	45

- Tính hệ số tương quan và nhận xét về sự phụ thuộc tính tuyến tính của X và Y (mạnh hay yếu? nghịch biến hay đồng biến?).
- Viết phương trình đường thẳng hồi quy tuyến tính để dự đoán thời gian hoàn thành dự án dựa trên số lượng dòng code. Ước tính thời gian hoàn thành một dự án có 4500 dòng code.

_____ Hết _____

Sinh viên được sử dụng tài liệu giấy và máy tính cầm tay.

Cán bộ coi thi không giải thích gì thêm.

Cán bộ ra đề

TRƯỞNG BỘ MÔN TOÁN - LÝ

CAO THANH BÌNH

ĐÁP ÁN VÀ BIỂU ĐIỂM

Câu 1: (1 điểm)

a. Ta có

$$P(X = 1, Y = -3) = 0,1 \neq P(X = 1)P(Y = -3) = (0,5) \cdot (0,4) = 0,2 \quad (0,25 \text{ điểm})$$

Suy ra, X, Y không độc lập. (0,25 điểm)

b. Tính

$$P(X > 1 | Y > 1) = \frac{P(X > 1 \cap Y > 1)}{P(Y > 1)} = \frac{0,2}{0,6} = 0,333 \quad (0,5 \text{ điểm})$$

Câu 2: (1,5 điểm)

a. Tìm $f_X(x)$. Với $0 \leq x \leq 2$

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dx dy = \int_0^2 \frac{3}{16} xy^2 dy = \frac{1}{16} xy^3 \Big|_0^2 = \frac{1}{2}x \quad (0,25 \text{ điểm})$$

Vậy

$$f_X(x) = \begin{cases} \frac{1}{2}x, & 0 \leq x \leq 2 \\ 0, & x \notin [0, 2] \end{cases} \quad (0,25 \text{ điểm})$$

b. Tính $P(Y \leq 1 | X = 1)$.

$$f_Y(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{3}{8}y^2, \text{ với } 0 \leq y \leq 2 \quad (0,25 \text{ điểm})$$

$$\begin{aligned} P(Y \leq 1 | X = 1) &= \int_0^1 f_Y(y|x) dy \\ &= \frac{1}{8}y^3 \Big|_0^1 = 0,125 \end{aligned} \quad (0,25 \text{ điểm})$$

c. Tính $P(X \leq Y)$.

$$\begin{aligned} P(X \leq Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^x f(x, y) dy dx \\ &= \int_0^2 \int_x^2 \frac{3}{16} xy^2 dy dx \end{aligned} \quad (0,25 \text{ điểm})$$

$$= \int_0^2 \frac{1}{16} x(8 - x^3) dx = \frac{15}{64} = 0,234375 \quad (0,25 \text{ điểm})$$

Câu 3: (1 điểm)

- Trung bình mẫu $\bar{x} = 57,13636$ (0,5 điểm)

- Độ lệch chuẩn mẫu $s = 7,158$ (hoặc độ lệch chuẩn mẫu hiệu chỉnh $s = 7,19$) (0,5 điểm)

Câu 4: (3 điểm) Gọi p tỉ lệ sinh viên sử dụng AI để hỗ trợ làm bài tập.

a. Ước lượng tỉ lệ

- Tỉ lệ mẫu $\hat{p} = \frac{300}{750} = 0,4$ và kích thước mẫu $n = 750$. (0,25 điểm)
- Độ tin cậy $1 - \alpha = 0,9$ suy ra $z_{\alpha/2} = 1,645$ (0,25 điểm)
- Độ chính xác (0,25 điểm)

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0,029$$

- Khoảng tin cậy của p là $(0,371; 0,429)$ (0,25 điểm)

b. Xác định kích thước mẫu

- Độ tin cậy $1 - \alpha = 0,95$ suy ra $z_{\alpha/2} = 1,96$ (0,25 điểm)
- Độ sai số tối đa (0,25 điểm)

$$\varepsilon = z_{\alpha/2} \frac{1}{2\sqrt{n}} \leq 0,02 \text{ hay } n \geq 49^2 = 2041$$

- Cần điều tra ít nhất 2041 sinh viên. (0,25 điểm)

c. Kiểm định giả thuyết

- Xét giả thuyết $H_0 : p = 0,38$ và đối thuyết $H_1 : p > 0,38$ (0,25 điểm)
- Tỉ lệ mẫu $\hat{p} = \frac{300}{750} = 0,4$ và kích thước mẫu $n = 750$.
- Vì mức ý nghĩa $\alpha = 0,05$ và kiểm định một phía phải nên suy ra giá trị tới hạn $z_{\alpha} = 1,645$ (0,25 điểm)
- Giá trị kiểm định (0,25 điểm)

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = 1,128$$

- Vì $z < z_{\alpha}$ nên không bác bỏ H_0 . (0,25 điểm)
- Với mức ý nghĩa 5%, ta chưa có đủ bằng chứng để ủng hộ phát biểu của giảng viên. (0,25 điểm)

Câu 5. (2,5 điểm)

Đặt μ (ms) là thời gian phản hồi trung bình của máy chủ.

a. Ước lượng thời gian phản hồi trung bình của máy chủ.

- Trung bình mẫu $\bar{x} = 242,92$ và độ lệch chuẩn mẫu $s = 10,39$ (0,25 điểm)
- Độ tin cậy $1 - \alpha = 95\%$, bậc tự do bằng 19, suy ra $t_{0,025} = 2,093$ (0,25 điểm)
- Độ chính xác (0,25 điểm)

$$\varepsilon = t_{\alpha/2} \frac{s}{\sqrt{n}} = 4,86$$

- Khoảng tin cậy cần tìm (238,06; 247,78). (0,25 điểm)

b. Ta kiểm định trung bình

- Giả thuyết $H_0 : \mu = 250$ và đối thuyết $H_1 : \mu < 250$. (0,5 điểm)
- Trung bình mẫu $\bar{x} = 242,92$; kích thước mẫu $n = 20$ và độ lệch chuẩn mẫu $s = 10,39$.
- Vì mức ý nghĩa $\alpha = 0,05$ và kiểm định một phía bên trái nên suy ra giá trị tới hạn tương ứng với bậc tự do bằng 19 là $t_\alpha = 1,729$ (0,25 điểm)
- Giá trị kiểm định

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = -3,047 \quad (0,25 \text{ điểm})$$

- Vì $t < -t_\alpha$ nên bác bỏ H_0 . (0,25 điểm)
- Ta đủ bằng chứng để ủng hộ phát biểu thời gian phản hồi trung bình của máy chủ giảm so với năm 2020. (0,25 điểm)

Câu 6. (1 điểm)

- a. Hệ số tương quan $r = 0,9987$ (0,25 điểm)
 Tương quan tuyến tính mạnh, đồng biến (0,25 điểm)
- b. Phương trình đường thẳng hồi quy tuyến tính (0,25 điểm)

$$y = 0,0096x + 3,3462$$

Khi $x = 4500$ thì $y = 46,54$. Nếu dự án có số dòng code là 4500 thì thời gian hoàn thành là 47 ngày (hoặc 46,5 ngày). (0,25 điểm)

_____ Hết _____